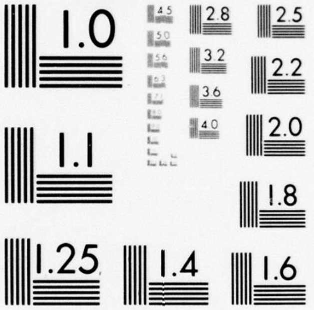
AD-A061 950 BROWN UNIV PROVIDENCE R I LEFSCHETZ CENTER FOR DYNAM--ETC F/G 12/1
JUMP-DIFFUSION APPROXIMATIONS FOR ORDINARY DIFFERENTIAL EQUATIO--ETC(U)
SEP 78 H J KUSHNER N00014-76-C-0279
UNCLASSIFIED LCDS-78-1 NL

1 OF 1
AD
A061950

END
DATE
FILMED
2-79
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF

Lefschetz Center for Dynamical Systems

ERRATA FOR

Jump-Diffusion Approximations for Ordinary Differential Equations with Wide-Band Random Right Hand Sides

H.J. Kushner

Page 7: Change (A3) to $\int_0^\infty \rho^{1/2}(t)dt < \infty$.

Delete the line below (4.1).

Page 8: Insert in footnote before its last sentence:

> The strong mixing (A3) also implies that for each bounded x-set there is a constant K such that
>
> $$|E\{F(x,y_{t+s})F'(x,y_{t+s+u})|y_a, a \le t\}$$
> $$- EF(x,y_{t+s})F'(x,y_{t+s+u})| \le K(\rho_s\rho_u)^{1/2}.$$

LEVEL H

(12)

(6)

JUMP-DIFFUSION APPROXIMATIONS FOR ORDINARY DIFFERENTIAL EQUATIONS WITH WIDE-BAND RANDOM RIGHT HAND SIDES[+]

by

(10) Harold J. Kushner

Divisions of Applied Mathematics and Engineering
Brown University
Providence, Rhode Island 02912

(15) N∅∅14-76-C-∅279,
✓AFOSR-76-3∅63

(12) 4∅p.

(11) Sep~~tember, 1~~978

(14) LCDS-78-1

Abstract

Let $y(\cdot)$ be a stationary mixing process and $J^{\varepsilon}(\cdot)$ an approximation to a random impulsive process. Kurtz's [1] results on approximation of a general semigroup by a Markov semigroup are used to prove (weak and a similar type of) convergence of the solutions to (1.1) and (1.2) to jumping diffusions. Previous results are generalized in various ways. The case of unbounded $y(\cdot)$ is also treated as is the combined jump-diffusion case. Also, a limit theorem for an integral with respect to "approximate white noise" in terms of an Itô integral is given. The method has the advantages of generality and relative ease of use.

superscript epsilon

401 834

[+] This research was partially supported by the Air Force Office of Scientific Research under AF-AFOSR 76-3063, the National Science Foundation under NSF-Eng73-03846-A-03 and by the Office of Naval Research under N0014-76-C-0279-P0002.

78 11 15 006

mt

1. Introduction

In [1], Kurtz gave some fairly general semigroup methods for showing convergence of a sequence of non-Markov process to a Markov process, either in the sense of weak convergence or in the sense of convergence of finite dimensional distributions. Let $y(\cdot)$ denote (a Euclidean space valued) right continuous strong mixing stationary process. For each $\varepsilon > 0$, define $y^{\varepsilon}(t) = y(t/\varepsilon^2)$, and for suitable F,G, define the process $x^{\varepsilon}(\cdot)$ by

$$\dot{x}_t^{\varepsilon} = \frac{F(x_t^{\varepsilon},y_t^{\varepsilon})}{\varepsilon} + G(x_t^{\varepsilon},y_t^{\varepsilon}), \quad x_0^{\varepsilon} = x_0 \in R^m, \quad y(t) \in R^{m'}. \tag{1.1}$$

Khazminskii [2], Papanicolaou [3], Papanicolaou and Kohler [4] and Blankenship and Papanicolaou [5] have all treated the problem of weak convergence of $x^{\varepsilon}(\cdot)$ to a diffusion. The problem is, of course, closely related to the original problem of Wong and Zakai [6]. In this paper, Kurtz's results, (together with a technique exploited in references [3] and [5]) will be used to get similar types of results under conditions which are weaker.[+] The method of proof has the great advantage of being quite straightforward and easy to use for both the diffusion and jump-diffusion cases.

We also treat limits of systems of the type

$$\dot{x}_t^{\varepsilon} = \frac{F(x_t^{\varepsilon},y_t^{\varepsilon})}{\varepsilon} + G(x_t^{\varepsilon},y_t^{\varepsilon}) + \sum_i H_i(x_t^{\varepsilon})J_{i,t}^{\varepsilon}, \tag{1.2}$$

where $\int_0^t J_{i,s}^{\varepsilon} ds$ is an approximation to a pure jump process, and we obtain a limit which is a jumping diffusion.

[+] References [2] and [3] allowed F and G to depend also on t. At the expense of extra detail, this case could be handled by our method.

Sections 2 and 3 recapitulate Kurtz's method of proving convergence of finite dimensional distributions and tightness, resp. The results are recapitulated partly for the sake of self-containment, and partly to state the precise form in which they are to be used. Section 4 states the assumptions used in Section 5, which gives the result for limits of (1.1) when $y(\cdot)$ is bounded, a restriction also used in the past references. Theorem 3 gives a result which is useful in approximating stochastic integrals with respect to a Wiener process by ordinary integrals. Such results are needed for identification and related problems (see Balakrishnan [8], [9]). A result on convergence of finite dimensional distributions for unbounded $y(\cdot)$ is given in Section 6, and tightness for unbounded $y(\cdot)$ is proved in Section 7. Section 8 deals with the relatively simple case where the input is an approximation to a random impulse process, and (1.2), an approximation to a jumping diffusion, is treated in Section 9. The result in Sections 6 and 7 cover the much used case where $y(\cdot)$ is a Gaussian diffusion.

A method similar to that of Section 5 is outlined in Section 4, [3], for the problem of showing convergence of finite dimensional distributions for the bounded $y(\cdot)$ case. The results there are not in a particularly usable form, and actually require more smoothness of F and G than needed here since partial differential equation methods are ultimately used there. Here, we do not need to solve or even to approximate solutions of partial differential equations, but merely to check the action of certain operators on smooth functions.

2. Convergence of Finite Dimensional Distributions*

Let $(\Omega, P, \mathscr{F})$ denote a probability space, $\{\mathscr{F}_t\}$ a nondecreasing sequence of sub σ-algebras of $\mathscr{F}$, $\mathscr{L}$ the space of progressively measurable real valued processes f on $[0,\infty)$, adapted to $\{\mathscr{F}_t\}$ and such that $\sup_t E|f(t)| < \infty$. Let f_n and f be in $\mathscr{L}$. Define the limit "p-lim" by p-lim $f_n = f$ iff $\sup_n \sup_t E|f_n(t)| < \infty$ and $E|f_n(t) - f(t)| \to 0$ for each t. For each $s > 0$, define the operator $\mathscr{T}(s): \mathscr{L} \to \mathscr{L}$ by $\mathscr{T}(s)f$ = function in $\mathscr{L}$ whose value at t is the random variable $E_{\mathscr{F}_t} f(t+s)$. There is a version which is progressively measurable ([1], Appendix) and we always assume that this is the one which is used. The $\mathscr{T}(s)$, $s \geq 0$, are a semigroup of linear operators on $\mathscr{L}$. Let $\hat{\mathscr{L}}_0$ denote the subspace of $\mathscr{L}$ of p-right continuous functions. If the limit $\text{p-lim}_{s\to 0} [\frac{1}{s}(\mathscr{T}(s)f-f)]$ exists and is in $\hat{\mathscr{L}}_0$, we call it $\hat{A}f$ and say that $f \in \mathscr{D}(\hat{A})$. The operators $\mathscr{T}(s)$ and $\hat{A}$ are analogous to the semigroup and weak infinitesimal operator of a Markov process. Among the properties to be used later is ([1], equation (1.9))

$$\mathscr{T}(s)f - f = \int_0^s \mathscr{T}(\tau)\hat{A}f\,d\tau, \quad f \in \mathscr{D}(\hat{A}), \tag{2.1a}$$

or, equivalently

---

*From [1], with slightly altered terminology. Sometimes we write f_t and sometimes $f(t)$ for the value of a process f at time t.

$$E_{\mathcal{F}_t} f(t+s) - f(t) = \int_0^s E_{\mathcal{F}_t} \hat{A}f(t+\tau)d\tau, \text{ for each } t \geq 0. \qquad (2.1b)$$

If, for some process $Z^\varepsilon(\cdot)$, $\mathcal{F}_t = \sigma(Z^\varepsilon_s, s \leq t)$, we may write* $\mathcal{F}^\varepsilon_t$, T^ε_t and $\hat{A}^\varepsilon$ for $\mathcal{F}_t$, $\mathcal{T}(t)$ and $\hat{A}$, resp. Let C_o and C^i_o denote the spaces of real valued functions on R^m which vanish at ∞ and which are continuous and which have continuous partial derivatives up to order i (and which also vanish at infinity), resp. Let $\hat{C}$ and $\hat{C}^i$ denote the sets of these functions which have compact support.

The following Theorem (a specialization of [1], Theorem 3.11) is our main tool for dealing with (1.1). Henceforth, unless otherwise mentioned, $\varepsilon \to 0$ replaces $n \to \infty$ in p-lim.

Theorem 1. Let $Z^\varepsilon(\cdot) = x^\varepsilon(\cdot), y^\varepsilon(\cdot)$, $\varepsilon > 0$, denote a sequence of $R^{m+m'}$ valued right continuous processes, $x(\cdot)$ a (R^m-valued) Markov process with semigroup T_t mapping C_o into C_o and which is strongly continuous (sup norm) on C_o. For some $\lambda > 0$ and dense set D in C_o, let Range $(\lambda - A|_D)$ be dense in C_o (sup norm, A = infinitesimal operator of $x(\cdot)$). Suppose that, for each $f \in D$, there is a sequence $\{f^\varepsilon\}$ of progressively measurable functions adapted to $\{\mathcal{F}^\varepsilon_t\}$ and such that

$$\text{p-lim}[f^\varepsilon - f(x^\varepsilon(\cdot))] = 0 \qquad (2.2)$$

$$\text{p-lim}[\hat{A}^\varepsilon f^\varepsilon - Af(x^\varepsilon(\cdot))] = 0. \qquad (2.3)$$

Then, if $x^\varepsilon(0) \to x(0)$ weakly, the finite dimensional distributions of $x^\varepsilon(\cdot)$ converge to those of $x(\cdot)$.

---

[*] The σ-algebras will often be completed, but the same notation will be used.

Equations (2.2) and (2.3) are equivalent to (the limits are taken for each t as $\varepsilon \to 0$)

$$\sup_{\varepsilon,t} E|f^{\varepsilon}(t) - f(x^{\varepsilon}(t))| < \infty, \quad E|f^{\varepsilon}(t) - f(x^{\varepsilon}(t))| \to 0 \tag{2.2'}$$

$$\sup_{\varepsilon,t} E|\hat{A}^{\varepsilon} f^{\varepsilon}(t) - Af(x^{\varepsilon}(t))| < \infty, \quad E|\hat{A}^{\varepsilon} f^{\varepsilon}(t) - Af(x^{\varepsilon}(t))| \to 0. \tag{2.3'}$$

3. Tightness

Let $y(\cdot)$, $y^{\varepsilon}(\cdot)$, $x^{\varepsilon}(\cdot)$ denote the functions in the model (1.1) or (1.2). *Let* $\mathcal{F}_t$ and $\mathcal{F}_t^{\varepsilon}$ *denote the (completed)* $\sigma(y_s; s \le t)$ and $\sigma(y_s^{\varepsilon}, s \le t)$. *Write* E_t *and* E_t^{ε} *for* $E_{\mathcal{F}_t}$ *and* $E_{\mathcal{F}_t^{\varepsilon}}$, *resp.*

Again, we describe results from [1]. Let $D^m[0,\infty)$ denote the space of R^m valued functions on $[0,\infty)$ which are right continuous on $[0,\infty)$ and have left hand limits on $(0,\infty)$. Note that $x^{\varepsilon}(\cdot) \in D^m[0,\infty)$ w.p.1. Suppose that the finite dimensional distributions of $x^{\varepsilon}(\cdot)$ converge to those of a process $x(\cdot)$, where $x(\cdot)$ has paths in $D^m[0,\infty)$ w.p.1. Then, as noted in [1], bottom of page 628, $\{x^{\varepsilon}(\cdot)\}$ is tight in $D^m[0,\infty)$ if $\{f(x^{\varepsilon}(\cdot))\}$ is tight in $D[0,\infty)$ for each $f \in \hat{C}$. ($\hat{C}$ is used there, but it can be replaced by[+] $\hat{C}^3$.) It follows from [1], Theorem 4.20, that $\{f(x^{\varepsilon}(\cdot))\}$ is tight in $D[0,\infty)$ if $x^{\varepsilon} \to x_0$ weakly and if, for each real $T > 0$, there is a random variable $\gamma_{\varepsilon}(\delta)$ such that

$$E_t^{\varepsilon}\gamma_{\varepsilon}(\delta) \ge E_t^{\varepsilon} \min\{1, [f(x_{t+u}^{\varepsilon}) - f(x_t^{\varepsilon})]^2\}, \tag{3.1}$$

for all $0 \le t \le T$, $0 \le u \le \delta \le 1$, and

[+] or by any set of functions dense in $\hat{C}$ in the sup norm.

$$\lim_{\delta\to 0} \overline{\lim_{\varepsilon\to 0}}\, E\, \gamma_\varepsilon(\delta) = 0. \tag{3.2}$$

In [1], p. 629, Kurtz suggests a method of getting the $\gamma_\varepsilon(\delta)$. This method is developed in Lemma 1 and is used in the sequel. The f^ε below will be obtained in the same manner as we will obtain the f^ε of Theorem 1. We have $(||f|| = \sup_x |f(x)|)$

$$\begin{aligned} E^\varepsilon_t[f(x^\varepsilon_{t+u}) - f(x^\varepsilon_t)]^2 &\le 2||f||\, \left|E^\varepsilon_t f(x^\varepsilon_{t+u}) - f(x^\varepsilon_t)\right| \\ &\quad + |E^\varepsilon_t f^2(x^\varepsilon_{t+u}) - f^2(x^\varepsilon_t)|. \end{aligned} \tag{3.3}$$

Lemma 1. *Let* $f \in \hat{C}^3$, *and let there be a sequence* $\{f^\varepsilon\}$ *in* $\mathcal{L}$, *where* $(f^\varepsilon)^i \in \mathcal{D}(\hat{A}^\varepsilon)$, $i = 1,2$, *and such that, for each real* $T > 0$ *there is a random variable* M *such that*

$$\sup_{t\le T} |f^\varepsilon(t) - f(x^\varepsilon_t)| \to 0 \text{ w.p.1, as } \varepsilon \to 0 \tag{3.4}$$

$$\sup_{\varepsilon>0,\, t\le T} |\hat{A}^\varepsilon(f^\varepsilon(t))^i| \le M, \text{ w.p.1}, \quad i = 1,2. \tag{3.5}$$

Then $\{f(x^\varepsilon(\cdot))\}$ *is tight in* $D[0,\infty)$.

Proof. By (2.1)

$$E^\varepsilon_t(f^\varepsilon(t+u))^i - (f^\varepsilon(t))^i = \int_0^u E^\varepsilon_t \hat{A}^\varepsilon (f^\varepsilon(t+\tau))^i d\tau,$$

from which (3.3) to (3.5) yield a $\gamma_\varepsilon(\delta)$ of the form (for the interval $[0,T]$) $\max\{1, \mathcal{U}\}$ where

$$\mathcal{Y} = M_1[\delta + \sum_{i=1}^{2} \sup_{t \le T} |(f^\varepsilon(t))^i - f^i(x_t^\varepsilon)|,$$

where M_1 is a random variable. Q.E.D.

4. *Assumptions for Model* (1.1); *Bounded* $y(\cdot)$.

(A1) $F(\cdot,\cdot)$ *and* $G(\cdot,\cdot)$ *are continuous, and the first partial and first and second x-partial derivatives of* G *and* F, *resp., are continuous.*

(A2) *There is a constant* M *such that*

$$|F(x,y)| + |G(x,y)| \le M(1+|x|).$$

(A1) and (A2) assure the global existence of solutions to (1.1).

(A3) $y(\cdot)$ *is a right continuous, bounded stationary process and* $EF(x,y_s) = 0$, *each* x.

There is a measurable function $\rho(\cdot)$ *such that*

$$\sup_{B_i,t} |P(B_2|B_1) - P(B_2)| \le \rho(\tau),$$

where $B_1 \in \sigma(y_u, u \le t)$, $B_2 \in \sigma(y_u, u \ge t + \tau)$. *Let*

$$\int_0^\infty \rho(t)dt < \infty. \tag{4.1}$$

(4.2) is weaker than the conditions on the mixing rates in [4].

Define the operator A on $\hat{C}^2$ by (the subscript x denotes gradient)

$$Af(x) = EG'(x,y_s)f_x(x) + \int_0^\infty d\tau \, EF'(x,y_s)(F'(x,y_{s+\tau})f_x(x))_x . \quad (4.2)$$

$$\equiv \sum_i b_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i,j} a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} .$$

By (A1) and (A3), the integrals on the right exist. In fact, the improper Lebesgue integral is absolutely convergent (i.e. $\int_0^T |E(\ )| d\tau$ converges) uniformly in x*. Furthermore, if $f \in \hat{C}^3$ then $Af(x)$ is continuously differentiable in x, and the gradient of $Af(x)$ is the function obtained by simply replacing the argument of E in (4.2) by its x-gradient. If this is done, then the improper Lebesgue integral still is absolutely convergent, uniformly in x.

(A4) A *is the restriction to* $\hat{C}^2$ *of the strong infinitesimal operator of a strong Markov conservative (no finite escape time) diffusion process* $x(\cdot)$, *with semigroup* T_t. T_t *maps* C_o *into* C_o *and is strongly continuous on* C_o.

(A5) $\{\mathcal{F}_t\}$ *is right continuous. I.e.,* $\mathcal{F}_t = \bigcap_{\delta>0} \mathcal{F}_{t+\delta}$, *each* $t \geq 0$.

(A6) *The set* $(\lambda-A)\hat{C}^2 \equiv \{g: g = (\lambda-A)f, f \in \hat{C}^3\}$ *is dense in* C_o *for some* $\lambda > 0$.

---

*This follows readily from the strong mixing. It is also a consequence of Billingsley [7], p. 170, by using $EF(x,y_s) \equiv 0$, (4.1) and the fact that the functions have bounded support (let $r \to 1$, $s \to \infty$ in [7], equation (20.23), with a proper identification of ξ, η there). We will use this and similar facts frequently in the sequel.

Remark on (A5). Let $f \in \hat{C}$ and let $\bar{f}$ denote either Ff, Gf of any of the g_i or f_i introduced below. Condition (A5) is introduced only because we want $E^{\varepsilon}_{t+s}\bar{f}(x^{\varepsilon}_{t+s}, t+s+u)$ to converge to $E^{\varepsilon}_{t}\bar{f}(x^{\varepsilon}_{t}, t+u)$ in probability as $s \downarrow 0$. Many of the calculations in Theorems 2 and 4 involve this type of right continuity together with uniform integrability.

Remark on (A6). Some condition such as (A6) is required for use of Theorem 1. Let A_c denote the strong infinitesimal operator of T_t acting on C_o. Then (A6) is equivalent to A_c being the closure of the operator A (of (4.2) acting on $\hat{C}^2$, or on $\hat{C}^3$, since $\hat{C}^3$ is dense in $\hat{C}^2$ in the norm $||f||_2 = \sup_x(|f(x)| + |f_x(x)| + |f_{xx}(x)|))$. This condition does not seem to be particularly restrictive. We only remark that it holds in the two extreme cases: (1) the $b_i(\cdot)$ and $a_{ij}(\cdot)$ in (4.2) are bounded, satisfy a uniform Hölder condition and $a(\cdot)$ is uniformly positive definite; (2) where T_t maps C^2_o into C^2_o and is strongly continuous on C^2_o with respect to the norm $||f||_2$ defined above. (The same remarks were made by Kurtz [1], p. 632.)

In case 2, we can actually consider T_t as acting on the Banach space C^2_o with norm $||f||_2$, and where $f(x)$ and its first and second derivatives go to zero as $|x| \to \infty$, and modify (A4) accordingly. In this case, the closure of the operator A (the domain of A is $\hat{C}^2$ here) is just the strong infinitesimal operator (of T_t) acting on its domain in C^2_o. Suppose that there is a matrix valued $\sigma(\cdot)$ such that $\sigma(x)\sigma'(x) = a(x)$, that (A4) holds (as modified above) and that $b_i(\cdot)$, $\sigma_{ij}(\cdot)$ are locally Lipschitz. Then (see remarks below on bounding), it is enough to prove Theorem 2 under the additional condition that $b_i(\cdot)$, $\sigma_{ij}(\cdot)$ are bounded and, for arbitrary N, arbitrarily smooth out of the sphere S_N of center 0 and radius N. Assume that $b_i(\cdot)$ and $\sigma_{ij}(\cdot)$ have continuous first and second derivatives. Then by the above remark on bounding, we can assume that the coefficients are bounded on R^m. Then (Gikhman and Skorokhod [13], Chapter 8.4), case (2) above holds. The conditions imposed are weaker than those in [4] when F and G do not depend on t.

Remark on bounding the coefficients. Suppose $a(x) = \sigma(x)\sigma'(x)$, where $b_i(\cdot)$ and $\sigma_{ij}(\cdot)$ are locally Lipschitz continuous and $x(\cdot)$, the diffusion with generator A, is conservative. Define an N-truncation as follows. Let $b_i^N(\cdot)$, $\sigma_{ij}^N(\cdot)$ equal $b_i(\cdot)$, $\sigma_{ij}(\cdot)$, resp., in S_N, be bounded on R^m, have bounded derivatives of any desired order in the complement of S_{2N} and be at least as smooth in $S_{2N} - S_N$ as $b_i(\cdot)$, $\sigma_{ij}(\cdot)$ are. Then the Itô process $x^N(\cdot)$ with coefficients $b_i^N(\cdot)$, $\sigma_{ij}^N(\cdot)$ is called an N-truncation of $x(\cdot)$ if the $\tilde{b}^N(\cdot)$ and $\tilde{\sigma}^N(\cdot)$ can be obtained by a modification of F and G in $R^m - S_N$. N-truncations always exist since we can multiply F and G by suitable smooth real valued functions $m_f(x)$ and $m_g(x)$, resp., which equal unity in S_N. We only need prove Theorem 2 and verify (A4) and (A6) for some N-truncation for each N.

The proof of the assertion will be omitted. It is essentially a note that the parts of $x^N(\cdot)$ and $x(\cdot)$ up to the first escape from S_N are equal, and that the probability of escape from S_N on an interval $[0,T]$ goes to zero as $N \to \infty$, for each fixed x_0. Here, we suppose that $x(\cdot)$ are defined with respect to the same Wiener process.

5. Proof of Weak Convergence; Bounded $y(\cdot)$ and (1.1)

The main job in using Theorem 1 is to get f^ε when f is given. To do this, we use an idea exploited for a similar purpose in Section 3 of [5] and in Section 4 of [3]. We look for functions of the form* $f^{\varepsilon}(x,t) = f(x) + \varepsilon f_1^{\varepsilon}(x,t) + \varepsilon^2 f_2^{\varepsilon}(x,t)$, and define $f^{\varepsilon}(x_t^{\varepsilon},t) = f^{\varepsilon}(t)$. Define operators $\hat{A}_x^{\varepsilon}$ and $\hat{A}_y^{\varepsilon}$ as follows. Let $g(x,t)$ be smooth as a function of x and such that $g(x,t)$ is a function of y_s^{ε}, $s \le t$, for each x. At $x = x_t^{\varepsilon}$, $y = y_t^{\varepsilon}$, let

$$\hat{A}_x^{\varepsilon} g(x,t) = g_x'(x,t)\left[\frac{F(x,y)}{\varepsilon} + G(x,y)\right]; \qquad (5.1)$$

i.e., $\hat{A}_x^{\varepsilon}$ is $\hat{A}^{\varepsilon}$, but acting on $g(x_t^{\varepsilon},t)$ considered only a function of its first argument. Let $\hat{A}_y^{\varepsilon} g(x,t)$ be $\hat{A}^{\varepsilon} g(x,t)$, but where g is considered to be a function of its second argument only. Assume for the moment that $\hat{A}^{\varepsilon} = \hat{A}_x^{\varepsilon} + \hat{A}_y^{\varepsilon}$. Then, in order to use Theorem 1, we apply $\hat{A}^{\varepsilon}$ to $f^{\varepsilon}(x,t)$, and insist that

$$\begin{aligned}
&[f_x(x) + \varepsilon f_{1,x}^{\varepsilon}(x,t) + \varepsilon^2 f_{2,x}^{\varepsilon}(x,t)]'\left[\frac{F(x,t)}{\varepsilon} + G(x,t)\right] + \\
&\qquad + \hat{A}_y^{\varepsilon}(\varepsilon f_1^{\varepsilon}(x,t) + \varepsilon^2 f_2^{\varepsilon}(x,t)) \qquad (5.1') \\
&= Af(x,t) + O(\varepsilon),
\end{aligned}$$

where equation (5.1') must determine both the operator A and equations yielding the $f_i^{\varepsilon}(x,t)$. In Theorem 2, we merely write down formulas for the $f_i^{\varepsilon}(x,t)$ and verify the conditions of Theorem 1.

Theorem 2. *Under* (A1) *to* (A6), $x^{\varepsilon}(\cdot)$ *converges weakly in* $D^m[0,\infty)$ *to the diffusion* $x(\cdot)$, *with initial condition* x_0.

Proof. First (Parts 1 to 3) we prove convergence of finite dimensional distributions, using Theorem 1. In Part 4 tightness is proved, via Lemma 1. Henceforth, f is a fixed element of $\hat{C}^3$. Since $\hat{C}^3$ is dense in $\hat{C}^2$ in the norm $||f||_2$, it is enough to work with $\hat{C}^3$.

---

*For each x and t, $f_i^{\varepsilon}(x,t)$ will be a function of y_s^{ε}, $s \le t$. The discussion in this paragraph is purely formal.

Part 1. $f_1^\varepsilon(x,t)$ is defined to be a solution (suggested by (5.1')) to $\hat{A}_y^\varepsilon f_1^\varepsilon(x,t) = -g_1(x,y_t^\varepsilon) \equiv -F'(x,y_t^\varepsilon)f_x(x)$. More precisely, define $f_1^\varepsilon(t) = f_1^\varepsilon(x_t^\varepsilon,t)$, where

$$f_1^\varepsilon(x,t) = \int_0^\infty E_t^\varepsilon g_1(x,y(\frac{t}{\varepsilon^2} + s))ds \tag{5.2}$$

$$= \frac{1}{\varepsilon^2}\int_0^\infty E_t^\varepsilon g_1(x,y_{t+s}^\varepsilon)\,ds$$

(both forms will be used). The improper Lebesgue integral exists and is bounded and absolutely convergent, uniformly in ω,x and in t in bounded sets, by the strong mixing (A3), and the facts that $EF(x,y_s) \equiv 0$ and that g_1 has compact x-support. Furthermore, there are versions of $f_1^\varepsilon(x,t)$ and $f_1^\varepsilon(x_t^\varepsilon,t)$ which are progressively measurable.

We next show that $f_1^\varepsilon(t) \in \mathscr{D}(\hat{A}^\varepsilon)$. We have

$$\begin{aligned}\hat{A}^\varepsilon f_\varepsilon^1(t) &= \underset{\delta\to 0}{\text{p-lim}}\,[E_t^\varepsilon f_1^\varepsilon(x_{t+\delta}^\varepsilon,t+\delta) - f_1^\varepsilon(x_t^\varepsilon,t)]/\delta \qquad (5.3)\\ &= \underset{\delta\to 0}{\text{p-lim}}\,[E_t^\varepsilon\{f_1^\varepsilon(x_{t+\delta}^\varepsilon,t+\delta) - f_1^\varepsilon(x_t^\varepsilon,t+\delta)\}]/\delta\\ &\quad + \underset{\delta\to 0}{\text{p-lim}}\,[E_t^\varepsilon f_1^\varepsilon(x_t^\varepsilon,t+\delta) - f_1^\varepsilon(x_t^\varepsilon,t)]/\delta\end{aligned}$$

if the limits exist and are in $\hat{\mathscr{L}}_0$. It is easy to verify that the second limit exists, is in $\hat{\mathscr{L}}_0$ and equals $-g_1(x_t^\varepsilon,y_t^\varepsilon)/\varepsilon^2$. Now, $f_1^\varepsilon(x,t)$ is differentiable in x. Indeed,

$$f_{1,x}^\varepsilon(x,t) = \frac{1}{\varepsilon^2}\int_0^\infty E_t^\varepsilon g_{1,x}(x,y_{t+s}^\varepsilon)ds,$$

since $\int_0^T |E_t^\varepsilon g_{1,x}(y_{t+s}^\varepsilon)|ds$ converges uniformly in x, and in t in bounded sets, as $T \to \infty$. This fact together with the representation

$$E_t^\varepsilon[f_1^\varepsilon(x_{t+\delta}^\varepsilon, t+\delta) - f_1^\varepsilon(x_t^\varepsilon, t+\delta)]/\delta = \tag{5.4}$$

$$\frac{1}{\delta}\int_0^\delta E_t^\varepsilon[f_{1,x}^\varepsilon(x_{t+u}^\varepsilon, t+\delta)]'[\frac{F(x_{t+u}^\varepsilon, y_{t+u}^\varepsilon)}{\varepsilon} + G(x_{t+u}^\varepsilon, y_{t+u}^\varepsilon)]du$$

and the facts that $f_{1,x}^\varepsilon$ is right continuous in the mean at t, and that the integrand is zero out of a bounded x,y range, can be used to show that the first limit on the right side of (5.3) exists, is in $\hat{\mathcal{L}}_0$ and equals $(x = x_t^\varepsilon,\ y = y_t^\varepsilon)$

$$[f_{1,x}^\varepsilon(x,t)]'[F(x,y)/\varepsilon + G(x,y)] = \hat{A}_x^\varepsilon f_1^\varepsilon(x,t). \tag{5.5}$$

Part 2. With $x = x_t^\varepsilon$, we will define $f_2^\varepsilon(t) = f_2^\varepsilon(x,t)$, where $f_2^\varepsilon(x,t)$ is the formal solution to

$$\begin{aligned}\hat{A}_y^\varepsilon f_2^\varepsilon(x,t) &= -g_2(x,t) \\ &\equiv -[F'(x,y_t^\varepsilon)f_{1,x}^\varepsilon(x,t) + G'(x,y_t^\varepsilon)f_x(x) - Af(x)]\ ,\end{aligned} \tag{5.6}$$

where A is defined in (4.2). More precisely, define f_2 by

$$f_2^\varepsilon(x,t) = \frac{1}{\varepsilon^2}\int_0^\infty E_t^\varepsilon g_2(x,t+s)ds. \tag{5.7}$$

There are versions of $f_2^\varepsilon(x,t)$ and $f_2^\varepsilon(x_t^\varepsilon,t)$ which are progressively measurable. We now ignore the G terms, for the difficulty lies with the $Ff_{1,x}^\varepsilon$ term. The improper Lebesgue integral $\int |E_t^\varepsilon g_2(x,t+s)|ds$ of (5.7) converges absolutely, uniformly in ω, x

and in t in any bounded set, by the strong mixing property and the definition of A. (Indeed, this is the reason for the choice of A.) To see this, note that (5.7) (without the G-terms) equals

$$\frac{1}{\varepsilon^2}\int_0^\infty ds\{E_t^\varepsilon F'(x,y_{t+s}^\varepsilon)f_{1,x}^\varepsilon(x,t+s) - EF'(x,y_{t+s}^\varepsilon)f_{1,x}^\varepsilon(x,t+s)\}, \qquad (5.8)$$

where the average value of the coefficient of ds is zero (by stationarity and the definition of A), and use the strong mixing condition. In fact, using the change of variables $s/\varepsilon^2 \rightarrow s'$, it can be seen from (5.8) and the strong mixing and compact x-support of f_1^ε that $|f_2^\varepsilon(t)|$ is bounded w.p.1, uniformly in x and ω and in bounded t intervals. Furthermore, $g_2(x,t)$ is continuously differentiable in x. In fact, the convergence assertion in the sentence above (5.4) also holds for g_2 replacing g_1 and

$$f_{2,x}(x,t) = \frac{1}{\varepsilon^2}\int_0^\infty E_t^\varepsilon g_{2,x}(x,t+s)ds.$$

Expression (5.3) holds <u>with</u> f_2^ε <u>replacing</u> f_1^ε if the limits exist and are in $\hat{\mathscr{L}}_0$. Again, we readily verify that the second limit in (5.3) exists, is in $\hat{\mathscr{L}}_0$ and equals $-g_2(x,t)/\varepsilon^2$ $(x = x_t^\varepsilon)$. An argument similar to that leading to (5.5) yields that the first limit in (5.3) exists, is in $\hat{\mathscr{L}}_0$ and equals (5.5) with f_2^ε replacing f_1^ε.

<u>Part 3</u>. Now, we apply Theorem 1. Since

$$\sup_{t,\varepsilon>0} E(|f_1^\varepsilon(t)| + |f_2^\varepsilon(t)|) < \infty,$$

we have

$$\text{p-lim}[f^\varepsilon - f(x^\varepsilon(\cdot))] = 0.$$

Now, calculate $\hat{A}^\varepsilon f^\varepsilon$. By Parts 1 and 2, with $x = x_t^\varepsilon$, $y = y_t^\varepsilon$,

$$\hat{A}^\varepsilon f^\varepsilon(x,t) = \hat{A}^\varepsilon f(x) + \varepsilon \hat{A}^\varepsilon f_1^\varepsilon(x,t) + \varepsilon^2 \hat{A}^\varepsilon f_2^\varepsilon(x,t) \tag{5.9}$$

$$= [F(x,y)/\varepsilon + G(x,y)]' f_x(x)$$

$$+ \varepsilon[-F'(x,y) f_x(x)/\varepsilon^2 + (F(x,y)/\varepsilon + G(x,y))' f_{1,x}^\varepsilon(x,t)]$$

$$+ \varepsilon^2[-\frac{1}{\varepsilon^2}\{F'(x,y) f_{1,x}^\varepsilon(x,t) + G'(x,y) f_x(x) - Af(x)\}$$

$$+ (F(x,y)/\varepsilon + G(x,y))' f_{2,x}^\varepsilon(x,t))] =$$

$$= Af(x) + \varepsilon[G'(x,y) f_{1,x}^\varepsilon(x,t) + F'(x,y) f_{2,x}^\varepsilon(x,t)]$$

$$+ \varepsilon^2 G'(x,y) f_{2,x}^\varepsilon(x,t).$$

We now readily verify that $\text{p-lim}[\hat{A}^\varepsilon f^\varepsilon - Af(x^\varepsilon(\cdot))] = 0$. Since $x_0^\varepsilon = x_0$, Theorem 1 implies that the finite dimensional distributions converge.

Part 4. Tightness. For tightness, we use Lemma 1. Each $f \in \hat{C}$ can be approximated uniformly arbitrarily closely by an $f \in \hat{C}^3$. Thus, by the Lemma and discussion preceeding it, we only need prove that $\{f(x^\varepsilon(\cdot))\}$ is tight for each $f \in \hat{C}^3$. Let $f \in \hat{C}^3$ and construct $f_1^\varepsilon, f_2^\varepsilon$ exactly as the $f_1^\varepsilon, f_2^\varepsilon$ were constructed above, and define $f^\varepsilon(t)$ as above. Then f^ε and $(f^\varepsilon)^2$ are in $\mathcal{D}(\hat{A}^\varepsilon)$ and there is a constant M such that w.p.1

$$\sup_{\varepsilon>0,\omega,t} |\hat{A}^{\varepsilon}(f^{\varepsilon}(t))^{i}| \leq M, \quad i = 1,2,$$

$$\sup_{t}|f^{\varepsilon}(t) - f(x^{\varepsilon}(t))| \to 0, \text{ as } \varepsilon \to 0.$$

Thus, Lemma 1 implies that $\{f(x^{\varepsilon}(\cdot))\}$ is tight in $D[0,\infty)$, for each $f \in \hat{C}^{3}$; hence $\{x^{\varepsilon}(\cdot)\}$ is tight in $D^{m}[0,\infty)$. The tightness, together with the convergence of finite dimensional distributions implies weak convergence. Q.E.D.

<u>An approximation to an integral.</u> In problems where changes of measure via Girsanov-like transformations are involved, such as occur in some identification problems (Balakrishnan [8], [9]), we need to get limits of integrals such as $z_t^{\varepsilon} = \int_0^t q'(x_s^{\varepsilon})(y_s^{\varepsilon}/\varepsilon)ds$, as $\varepsilon \to 0$.

Let $\hat{C}^{i}$ denote the real valued functions on $R^{m+m'+1}$ with compact support, whose i^{th} partial derivatives are continuous. Let $Y_t^{\varepsilon} = \int_0^t (y_s^{\varepsilon}/\varepsilon)ds$ and $u^{\varepsilon} = (x^{\varepsilon},Y^{\varepsilon},z^{\varepsilon})$. Then u_t^{ε} is $R^{m+m'+1}$ valued and

$$\dot{u}^{\varepsilon} = \tilde{G}(x^{\varepsilon},y^{\varepsilon}) + \frac{1}{\varepsilon}\tilde{F}(x^{\varepsilon},y^{\varepsilon}),$$

where

$$\tilde{G}(x,y) = (G(x,y),0,0)$$
$$\tilde{F}(x,y) = (F(x,y),y,q'(x)y).$$

<u>The remarks below</u> (A6) <u>all pertain here also</u>.

Theorem 3. Let $q(\cdot)$ *satisfy the conditions on* $F(\cdot)$ *in* (A1)-(A2). *Assume* (A1) *to* (A6) *where* (A4) *and* (A6) *hold for the process* $u(\cdot)$ *and operator* $\bar{A}$ *defined on* $\hat{C}^2$ *by*

$$\bar{A}f(u) = E\tilde{G}'(x,y_0)f_u(u) + \int_0^\infty d\tau E\tilde{F}'(x,y_0)(\tilde{F}'(x,y_\tau)f_u(u))_u . \qquad (5.10)$$

Then $u^\varepsilon(\cdot)$ *converges weakly in* $D^{m+m'+1}[0,\infty)$ *to* $u(\cdot) = (x(\cdot), Y(\cdot), z(\cdot))$, $u_0 = (x_0,0,0)$, *a diffusion with generator* $\bar{A}$ *on* $\hat{C}^2$. *The process* Y_t *is a Brownian motion with covariance* $R = 2\int_0^\infty Ey_t y_0' d\tau$. *The limit* $z(\cdot)$ *has the Itô representation (the expectation is over* $y(\cdot)$) *in terms of the limits* $x(\cdot), Y(\cdot)$

$$dz = q'(x)dY + [\int_0^\infty EF'(x,y_0)(q'(x)y_\tau)_x d\tau]dt. \qquad (5.11)$$

(The last term in (5.11) is the so-called correction term of the limiting integral approximation.)

The proof of weak convergence of $u^\varepsilon(\cdot)$ to $u(\cdot)$ is simply an application of Theorem 2. Once the weak convergence is known, then the representation (5.11) is not hard to get, and we omit the details.

6. *Unbounded $y(\cdot)$ and (1.1).*

Our approach here will be only a little different from that in Section 5. In order to avoid conditions which look overly complicated, we specialize $F(x,y)$ and $G(x,y)$ to $F(x)y$ and $G(x)$, resp.

Assumptions. In this section, the convergence of finite dimensional distribution is proved, and tightness is treated in the next section. Owing to the unboundedness of $y(\cdot)$, it is convenient to artificially bound the F and G. We do this by dealing with a sequence of approximations to the original processes. The operator A is still defined by (4.2).

(B1) $F(\cdot)$ and its first and second order partial derivatives are continuous.

(B2) $G(\cdot)$ and its first order partial derivatives are continuous.

(B3) $\{\mathcal{F}_t\}$ (again completed) is right continuous, and so is the stationary process $y(\cdot)$, w.p.1. (See the remark concerning (A5).)

Define

$$v_t^\varepsilon = \frac{1}{\varepsilon^2}\int_0^\infty E_t^\varepsilon y_{t+s}^\varepsilon ds, \quad v_t = \int_0^\infty E_t y_{t+s} ds.$$

(B4) For some $\rho > 0$, $\sup_t E\left(\int_0^\infty |E_t y_{t+s}| ds\right)^{2+\rho} < \infty$.

Thus, v_t^ε and v_t are well-defined.

(B5) $E|y_t|^{2+\rho} < \infty$, some $\rho > 0$.

(B6) $\sup_t E\left(\int_0^\infty ds |E_t y_{t+s} v'_{t+s} - E\, y_{t+s} v'_{t+s}|\right)^{2+\rho} < \infty$, some $\rho > 0$.

Note that $Ey_{t+s}v'_{t+s} = Ey_{t+s}\int_0^\infty E_{t+s}y'_{t+s+u}du = \int_0^\infty Ey_t y'_{t+u}du$ and does not depend on t or s, and is well-defined by (B4), (B5).

Owing to the special form of F and G, there are locally Lipschitz $b(\cdot)$ and $\sigma(\cdot)$ such that (see (4.2)) $\sigma(\cdot)\sigma'(\cdot) = a(\cdot)$. See also the remarks on bounding below (A6). x^N and $x^{\varepsilon,N}(\cdot)$ denote the N-truncations of $x(\cdot)$ and $x^{\varepsilon}(\cdot)$.

(B7) For a sequence $N \to \infty$, there are N-truncations which satisfy (A4) and (A6).

Remark on assumptions (B3)-(B6). Let $w(\cdot)$ denote a vector Brownian motion, Q a matrix with eigenvalues in the open left half plane and let D and G be matrices. Define processes $Y(\cdot)$ and $y(\cdot)$ by

$$\begin{aligned} dY &= QY\,dt + Ddw, \\ y(\cdot) &= GY(\cdot). \end{aligned} \tag{6.1}$$

Then (B3) to (B6) hold. In this case, we can let $\mathcal{F}_t^{\varepsilon}$ and $\mathcal{F}_t$ measure Y_s, $s \le t/\varepsilon^2$, and Y_s, $s \le t$, resp., in all the foregoing. Then v_t is proportional to Y_t and $|E_t y_{t+s}| \le |Y_t| c_1 e^{-c_2 s}$, where the c_i are positive constants.

Theorem 4 deals with the convergence of finite dimensional distributions.

Theorem 4. Under (B1) to (B7) and the first sentence of (A4), the finite dimensional distributions of $x^{\varepsilon}(\cdot)$ converge to those of $x(\cdot)$ with initial condition x_0, as $\varepsilon \to 0$.

Proof. If the finite dimensional distributions of $x^{\varepsilon,N}(\cdot)$ converge to those of $x^N(\cdot)$ (initial condition x_0) as $\varepsilon \to 0$ for a sequence $N \to \infty$, then the conservative and the strong Markov properties of (A4), (B7) yield the theorem. So we only need prove convergence for a fixed N. Consequently, we may assume that F and G are bounded and drop the affixes N.

The details are very similar to those of Theorem 2, and we only make a few remarks. As before, define f_1^ε and f_2^ε by

$$f_1^\varepsilon(t) = f_1^\varepsilon(x_t^\varepsilon,t),\ f_2^\varepsilon(t) = f_2^\varepsilon(x_t^\varepsilon,t),$$

where $f_i^\varepsilon(x,t)$ is defined as in Theorem 2. These functions are no longer bounded, but still $\sup\limits_{t,\varepsilon>0} E|f_i^\varepsilon(t)| < \infty$. It is rather straightforward to verify (in fact, easier than in Theorem 2 owing to the special form of F(x,y) and G(x,y) here) via (B1) to (B6), that $f_i^\varepsilon \in \mathcal{D}(\hat{A}^\varepsilon)$ and* take the same values as in Theorem 2. Furthermore, the expectations of the absolute values of the coefficients of ε and ε^2 on the far right side of (5.9) are bounded, uniformly in ε. Also

$$\sup_{t,\varepsilon>0} E\{|\hat{A}^\varepsilon f_i^\varepsilon(t)| + |f_i^\varepsilon(t)|\} < \infty.$$

Thus,

---

*For example, to show that the expression for $\hat{A}^\varepsilon f_i^\varepsilon(t)$ is in $\hat{\mathcal{L}}_0$, we note that the compact support of f and (B4) to (B6) imply uniform integrability of the expression. This, together with (B3) yields p-right continuity.

$$\text{p-lim}[f^\varepsilon - f(x^\varepsilon(\cdot))] = 0 \quad \text{and} \quad \text{p-lim}[\hat{A}^\varepsilon f^\varepsilon - Af(x^\varepsilon(\cdot))] = 0,$$

from which the theorem follows, by Theorem 1. Q.E.D.

7. Tightness; Unbounded $y(\cdot)$

Owing to the unboundedness, it is more difficult to prove tightness via the method of Lemma 1. To avoid (what are at the moment) awkward conditions, we suppose that $y(\cdot)$ satisfies (6.1). Then the f^ε can be explicitly evaluated and the proof is easy.

Theorem 5. Assume (B1), (B2), (B7), the first sentence of (A4) and that $y(\cdot)$ satisfies (6.1). Then $\{x^\varepsilon(\cdot)\}$ is tight and converges weakly to $x(\cdot)$ as $\varepsilon \to 0$.

Remark. The *tightness* argument only uses the compact support of f, (B1)-(B2) and (6.1).

Proof. The method and notation of Lemma 1 and Theorem 2, Part 4, will be used here. We need to show that, for each T and each $f \in \hat{C}^3$

(i) $\lim\limits_{\varepsilon\to 0} \sup\limits_{t\le T} |f(x_t) - f^\varepsilon(t)| = 0$ w.p.1

(ii) $(f^\varepsilon)^2 \in \mathcal{D}(\hat{A}^\varepsilon)$

(iii) $\overline{\lim\limits_{\varepsilon\to 0}} \sup\limits_{t\le T} |\hat{A}^\varepsilon (f^\varepsilon(t))^j| < \infty$, $j = 1,2$, w.p.1.

In our case, there is a matrix C_0 such that

$$f_1^{\varepsilon}(x,t) = [F(x)C_0Y(t/\varepsilon^2)]'f_x(x)$$

$$g_2(x,t) = \{F(x)y(t/\varepsilon^2)\}'\{[F\ (x)C_0Y(t/\varepsilon^2)]'f_x(x)\}_x$$
$$+ G'(x)f_x(x) - Af(x).$$

$f_2^{\varepsilon}(x,t)$ is a quadratic form in the components of $Y(t/\varepsilon^2)$ where the coefficients are bounded differentiable functions of x with compact support. Also, as is readily verifiable, $(f^{\varepsilon}(t))^2 \in \mathcal{D}(\hat{A}^{\varepsilon})$, and $\hat{A}^{\varepsilon}(f^{\varepsilon}(t))^j, j = 1,2,$ have terms in powers of the components of $Y(t/\varepsilon^2)$ up to $2j+1$. Thus, to verify (i) and (iii) it is enough to verify that for each $T > 0$

$$\lim_{\varepsilon\to 0} \sup_{t\leq T} \varepsilon|Y(t/\varepsilon^2)|^3 = 0 \quad \text{w.p.1.} \tag{7.1}$$

Equation (7.1) holds since, for each $\alpha > 0$, the Gaussianess, stationarity, and special form (6.1) imply that there are finite w.p.1 ω-functions C_1 and C_2 such that $|Y(t)| \leq C_1t^{\alpha} + C_2$ for all t, w.p.1. Q.E.D.

8. An Approximate Jump Case; (1.2) With No $y^{\varepsilon}(\cdot)$ Term

Since the classical papers of Wong and Zakai [6], the problem of using Itô or other types of equations to approximately model processes which are the solutions to ordinary differential equations has received much attention; e.g., [3]-[5] and Sections 5 to 7 above.

Alternative approaches have been taken by McShane [10] and Sussmann [11] who sought either a theory of integration or a differential equation and a topology on the input functions so that the output is a continuous function of the input. The differential equations which were of the so-called Stratonovich form, which, in fact, are precisely Itô equations with appropriate dynamic terms.

Little has been done when the input is an approximation to an impulse (its integral is an approximation to a pure jump process). Marcus [12] has done some work along McShane's "belated integral" point of view. This work [12] has some interesting aspects, but also a number of shortcomings. The dynamics are rather special, being (in part) analytic functions. This is a disadvantage in any approximation theory, where robustness is a key word. Some heavy lie algebra machinery was used, and the form of the results tended to obscure the basic simplicity of the problem. Also, a very particular impulse approximation was used (piecewise constant). In this section, we take a simple minded but inherently natural and robust approach, using pathwise approximations and limits. The limits being either ordinary or stochastic differential equations with impulsive or jump inputs.

Let $\bar{N}_i(ds\times dy)$, $i = 1,\dots,k$, denote a sequence of scalar valued random measures and define $N_i(t) = \int\int_0^t \alpha\, \bar{N}_i(ds\times d\alpha)$, where $N_i(\cdot)$ is taken to be right continuous. The range of the jumps of $N_i(\cdot)$ is a bounded set R_1. Each $N_i(\cdot)$ is assumed to have a finite number of jumps on each bounded interval w.p.1, and the probability is zero that different $N_i(\cdot)$ have simultaneous jumps.

In this section, we deal with the equation

$$\dot{x}^\varepsilon_t = G(x^\varepsilon_t) + \sum_{i=1}^{k} H_i(x^\varepsilon_t)J^\varepsilon_i(t), \tag{8.1}$$

where the input $J^\varepsilon_i(t)$ is an "approximation" to the impulse $\dot{N}_i(t)$. Figure 1 illustrates some ways in which an actual integrated input $\int_0^t J^\varepsilon_i(s)ds$ might approximate an ideal integrated impulsive input $N_i(t)$. In the figure a jump of Y occurs at $t = t_0$. With approximation (1), $J_i(t) = Y/\varepsilon$ on $[t_0,t_0+\varepsilon]$. Define $K_i(x,y)$ to be such that $x + K_i(x,Y)$ is the solution to $\dot{x} = H_i(x)Y$ at $t = 1$, with $x_0 = x$.

It is convenient to start with the vector of *ideal* integrated inputs $N(\cdot) = \{N_i(\cdot), i \le k\}$ and to get the *actual* inputs J^ε_t from this, as indicated in Figure 1. When the parameter is ε, we work only with paths for which the interjump times of the vector $\{N_i(\cdot), i = 1,\dots,k\}$ are $\ge \varepsilon$. Obviously, this involves neglecting a set of paths whose probability goes to zero as $\varepsilon \to 0$, and it has no effect on the limiting process. Thus, for our "limit" results, only the case of one input need be treated, and unless noted otherwise set $k = 1$ and drop the indices i on J,N,H and K. Let $N(\cdot)$ jump Y_j at $t = \sigma_j$, its j^{th} jump time. Define $p^\varepsilon_t = J^{\dot{\varepsilon}}_t/Y_j$ on $[\sigma_j,\sigma_j+\varepsilon]$, and equal to zero out of $\bigcup_j [\sigma_j,\sigma_j+\varepsilon]$ and define $P^\varepsilon_t = \int_0^t p^\varepsilon_s ds$. Thus

$p^{\varepsilon}_{\sigma_j+\varepsilon} - p^{\varepsilon}_{\sigma_j} = 1$. Now (8.1) can be rewritten in the form

$$\begin{aligned} \dot{x}^{\varepsilon}_t &= G(x^{\varepsilon}_t) + H(x^{\varepsilon}_t)Y_j p^{\varepsilon}_t \quad \text{on} \quad [\sigma_j,\sigma_j+\varepsilon], \\ &= G(x^{\varepsilon}_t) \qquad\qquad\qquad \text{otherwise.} \end{aligned} \tag{8.2}$$

The value of $p^{\varepsilon}(\cdot)$ can depend on the jump time and size and on the state prior to the jump (and on the index i in case (8.1)).

Assumptions.

(C1) *The* G *and* H_i *are continuous.*

(C2) *There is one and only one solution to* $\dot{x} = G(x)$, *for each* x_0 *in* R^m; *for each* $T < \infty$, *the solution is bounded on* $[0,T]$ *uniformly in bounded* x_0 *sets.*

(C3) *For each* i, *there is one and only one solution on* $[0,1]$ *to* $\dot{w} = H_i(w)Y$ *for each* $Y \in R_i$ *and* $w_0 \in R^m$. *This solution is bounded, uniformly on bounded* (w_0,Y) *sets.*

(C4) *There are real numbers* $m_{\varepsilon} > 0$, $M_{\varepsilon} < \infty$, *such that* $m_{\varepsilon} \leq p^{\varepsilon}(t) \leq M_{\varepsilon}$ *on the* $[\sigma_j,\sigma_j+\varepsilon]$ *intervals and* $p^{\varepsilon}(\cdot)$ *is continuous.*

(C5) *Let* $0 < \mu^{\varepsilon}(t)$, *where* $\mu^{\varepsilon}(\cdot)$ *is bounded and continuous on* $[0,1]$ *and* $\int_0^1 \mu^{\varepsilon}(s)ds \leq \varepsilon$. *Define* $w^{\varepsilon}(\cdot)$ *by* $\dot{w}^{\varepsilon} = G(w)\mu^{\varepsilon}(t) + H_i(w)Y$. *For each* i, *let* $w^{\varepsilon}(\cdot)$ *exist and be bounded on* $[0,1]$ *uniformly on bounded* (Y,w^{ε}_0) *sets and in* $\{\varepsilon,\mu^{\varepsilon}(\cdot),\ \varepsilon \leq \varepsilon_0\}$ *for some* $\varepsilon_0 > 0$.

Remarks. (C2) implies that $x(\cdot)$ is continuously dependent on x_0, and (C3) implies that (for each i) $w(\cdot)$ is continuously dependent on w_0, Y. (C2), (C3) and (C5) are partially redundant, but it seemed easier to state the assumptions in this way.

Let $\dot{w} = H_i(w)Y$ and $w_0^\varepsilon \to w_0$. Then (C5) and (C3) imply that $|w^\varepsilon(t) - w(t)| \to 0$ as $\varepsilon \to 0$ uniformly on $[0,1]$ and on bounded (Y, w_0) sets, and in $\{\mu^\varepsilon(\cdot)\}$.

Theorem 6. *Assume* (C1) *to* (C5). *Let* $x(\cdot)$ *be defined by*

$$x_t = x_0 + \int_0^t G(x_s)ds + \sum_i \int_{R_i} K_i(x_{s^-},\alpha)\bar{N}_i(d\alpha \times ds). \tag{8.3}$$

Let $x_0^\varepsilon \equiv x_0$. *Then for each* $T < \infty$,

$$\sup_{t \in T_\varepsilon} |x_t^\varepsilon - x_t| \to 0 \text{ as } \varepsilon \to 0, \text{ w.p.1},$$

where $T_\varepsilon = [0,T] - \bigcup_j [\sigma_j, \sigma_j + \varepsilon]$.

If the dimension $m = 1$, *then* $x^\varepsilon(\cdot) \to x(\cdot)$ *in the Skorokhod topology* ([7], Section 14).

Remarks. (8.3) is the correct limit equation - the analog of the Wong-Zakai or Stratonovich equation for the modelling of the output of a system with approximate jump inputs. We do not quite have weak convergence or convergence in the Skorokhod metric for almost all paths, unless the components of the solution to $\dot{w} = H_i(w)Y$ are monotonic in $[0,1]$ for each Y and w_0. To see why, let $m \geq 2$. Then it is possible that $w_1(t)$, the first (or any other) component, behaves as in Figure 2; i.e., it does not approach its value at time 1 monotonically. Let $N_i(\cdot)$ jump Y at time σ, and suppose that there are no other jumps on the interval $[\sigma-\varepsilon,\sigma+\varepsilon]$. For small ε, the first component of the solution to (8.1) on $[\sigma,\sigma+\varepsilon]$ is essentially a compression of the $[0,1]$ segment in Figure 2 to the $[0,\varepsilon]$ interval, together with a shift of σ in the time origin. Owing to the non-monotonicity, we *cannot* have weak convergence - or pathwise convergence in the Skorokhod topology, since in the limit, as $\varepsilon \to 0$, the compression of the curve in Figure 2 gives a triple jump at time 0. Because of this, the assertions of Theorems 6 and 7 are a little awkward.

A main feature of the theorem is the robustness of the result; the limit does not depend on the precise form of the approximations $J_i^\varepsilon(\cdot)$. Obviously, the interpolation need not be over only an ε-interval.

Proof. We need treat only one jump and one H_i term, owing to the assumptions on $N_i(\cdot)$ and on the continuity with respect to parameters implied by (C2), (C3) and (C5)'. So return to (8.2) with σ_j set equal to zero.

We change the time scale. Define a monotone increasing function on $[0,\varepsilon]$, $\tau(t) = \tau$ by $d\tau/dt = p_t^\varepsilon$ or $\tau(t) = \int_0^t p_s^\varepsilon ds$. Thus, $\tau(\varepsilon) = 1$, and the inverse $t(\tau)$ exists by (C4). Define $z^{\varepsilon}(\tau) = x^{\varepsilon}(t(\tau))$. Then

$$\frac{dz^{\varepsilon}(\tau)}{d\tau} = G(z^{\varepsilon}(\tau))\mu^{\varepsilon}(\tau) + H(z^{\varepsilon}(\tau))Y, \quad \tau \leq 1, \tag{8.4}$$

where

$$\mu^{\varepsilon}(\tau) = [p^{\varepsilon}(t(\tau))]^{-1} = \frac{dt(\tau)}{d\tau},$$

$$z^{\varepsilon}(0) = x^{\varepsilon}(0).$$

Now $\int_0^1 \mu^{\varepsilon}(\tau)d\tau = \varepsilon$ and $\mu^{\varepsilon}(\cdot)$ satisfies the conditions in (C5). Let $x^{\varepsilon}(0) \to x(0)$ as $\varepsilon \to 0$. Then $x^{\varepsilon}(\varepsilon) \to K(x(0),Y) + x(0)$ as $\varepsilon \to 0$. This, together with the continuity conditions (C2), (C3), (C5) and a concatenation of the argument, implies the theorem. Q.E.D.

9. The Jump-Diffusion Case (1.2); Bounded y(·)

We now return to the full model (1.2). Owing to the possible non-weak convergence of $x^{\varepsilon}(\cdot) \to x(\cdot)$ in the pure jump case, due to the convergence problem at the jump times σ_j which was illustrated in connection with Figure 2, the combined jump-diffusion case will be treated by a 'piecing together' argument. Here $\bar{N}_i(\cdot)$, $i = 1,\ldots,k$, are independent Poisson random measures with rates $\lambda_i > 0$, jump distributions $D_i(\cdot)$ (with bounded support R_i) and are independent of $y(\cdot)$. With A defined by (4.2), define the operator A_J on $\hat{C}^2$ by

$$A_J f(x) = Af(x) + \sum_i \lambda_i \int [f(x+K_i(x,\alpha)) - f(x)]D_i(d\alpha). \quad (9.1))$$

Let $\bar{x}(\cdot)$ denote the jump-diffusion process whose infinitesimal operator on $\hat{C}^2$ is A_J. Except for the non-convergence problem at the jump points, $x^\varepsilon(\cdot)$ will essentially converge weakly to $\bar{x}(\cdot)$.

Set $\sigma_0 = 0$ and $\sigma_q = q^{th}$ jump of the vector valued process $N(\cdot) = \{N_i(\cdot),\ i \leq k\}$. Let the q^{th} jump be a jump of $N_{\ell_q}(\cdot)$ and have value Y_q. Define x_0^ε to be the solution of (1.1) with initial condition x_0 and let $x_q^\varepsilon(\cdot)(q \geq 1)$ be the solution to (1.1) with initial condition $x^\varepsilon(\sigma_q+\varepsilon)$, and where $y^\varepsilon(\sigma_q+\varepsilon+\cdot)$ is used in lieu of $y^\varepsilon(\cdot)$. Let $x_q(\cdot)$ denote the diffusion of Sections 4 and 5, with initial condition x_0 if $q = 0$, and $x_q(0) = x_{q-1}(\sigma_q-\sigma_{q-1}) +$ $+ K_{\ell_q}(x_{q-1}(\sigma_q-\sigma_{q-1}),Y_q)$ in general. We will need either (D1) or (D2) to replace (C5). Let $A_i(Y)$ denote the operator on $\hat{C}^1$ functions which is defined by $(H_i(w)Y)' \frac{\partial}{\partial w}$.

(D1) $\mu^\varepsilon(s)/\varepsilon$ *is bounded in* $s \leq 1$ *and* ε. *The functions* F, G *and* H_i *are bounded.*

(D2) $\mu^\varepsilon(s)/\varepsilon$ *is bounded in* $s \leq 1$ and ε. *Each* $A_i(Y)(Y \in R_i)$ *is the strong infinitesimal operator of a (degenerate) conservative Markov semigroup mapping (and strongly continuous on)* C_0 *into* C_0. *Also* $(\lambda - A_i(Y))\hat{C}^2$ *is dense in* C_0 *for each* $Y \in R_i$ *and* i, *and some* $\lambda > 0$ *(which can depend on* Y *and* i).

Theorem 7. *Assume* (A1) *to* (A6), (C1), (C4) *and either* (D1) *and* (C3) *or* (D2). *Then for each* N, $\{x_q^\varepsilon(\cdot),\ q \leq N\}$ *converges to*

$\{x_q(\cdot),\ q \le N\}$ <u>weakly in</u> $D^{mN}[0,\infty)$.

<u>Note</u>. The remarks below (A6) apply here also.

<u>Proof</u>. Owing to the independence of $N(\cdot)$ and $y(\cdot)$ and right continuity of $y(\cdot)$, $y(\sigma_j+\varepsilon+\cdot)$ has the properties of $y(\cdot)$. Due to this independence, the independent increments property of $N(\cdot)$ and the uniqueness and strong Markov property of the $x(\cdot)$ of Sections 4 and 5 we can use a "piecing together" method based on the following assertion: Let a component $N_\ell(\cdot)$ jump Y at $t = \sigma$, with no other jumps on $[\sigma-\varepsilon,\sigma+\varepsilon]$, and let $x^\varepsilon(\sigma) \equiv \tilde{x}^\varepsilon(\sigma) \to \tilde{x}(\sigma)$ weakly as $\varepsilon \to 0$ and define $\tilde{x}^\varepsilon(\cdot)$ for $t \in (\sigma,\sigma+\varepsilon]$ by

$$\dot{\tilde{x}}^\varepsilon_s = G(\tilde{x}^\varepsilon_s, y^\varepsilon_s) + \frac{1}{\varepsilon} F(\tilde{x}^\varepsilon_s, y^\varepsilon_s) + H_{\ell_1}(\tilde{x}^\varepsilon_s) p^\varepsilon_s Y. \tag{9.2}$$

Then (to be proved) $\tilde{x}^\varepsilon(\sigma+\varepsilon)$ converges weakly to $\tilde{x}(\sigma) + K_{\ell_1}(\tilde{x}(\sigma),Y)$, as $\varepsilon \to 0$. We will prove only the assertion, and the proof uses a combination of the ideas in Theorems 2 and 6. For notational simplicity, let $\sigma = 0$ and drop the index ℓ_1.

As in Theorem 6, change the time scale by defining $\tau(\cdot)$ on $[0,\varepsilon]$ and $w^\varepsilon(\cdot)$ by $d\tau(t)/dt = p^\varepsilon(t)$, $w^\varepsilon(\tau) = \tilde{x}^\varepsilon(t(\tau))$, where $t(\cdot)$ is the inverse of $\tau(\cdot)$. Then

$$\begin{aligned}\frac{dw^\varepsilon(\tau)}{d\tau} &= G(w^\varepsilon(\tau), y^\varepsilon(t(\tau)))\mu^\varepsilon(\tau) + \frac{\mu^\varepsilon(\tau)}{\varepsilon} F(w^\varepsilon(\tau), y^\varepsilon(t(\tau))) \\ &\quad + H(w^\varepsilon(\tau))Y, \qquad w^\varepsilon(0) = \tilde{x}^\varepsilon(0),\end{aligned} \tag{9.3}$$

where $\mu^{\varepsilon}(\tau) = [p^{\varepsilon}(t(\tau))]^{-1} = dt(\tau)/d\tau$. We need only show that, for fixed Y, $w^{\varepsilon}(1)$ converges weakly to $\tilde{x}(0) + K(\tilde{x}(0),Y)$, the value at $t = 1$ of the solution $w(\cdot)$ to $\dot{w} = H(w)Y$, $w(0) = \tilde{x}(0)$, $\sigma = 0$. We will actually prove the stronger result of weak convergence of $w^{\varepsilon}(\cdot)$ to $w(\cdot)$ in $D^m[0,1]$, for each fixed non-random Y.

First, the proof under (D2) will be given. Using the method of Theorem 2, let $f \in \hat{C}^2$ and set $\mu^{\varepsilon}(\tau) = 0$ and $t(\tau) = t(1) = \varepsilon$ for $\tau \geq 1$ and define $f_1(w,\tau)$ by $(\tau \leq 1)$

$$f_1^{\varepsilon}(w,\tau) = \frac{1}{\varepsilon^2}\int_0^{\infty} E^{\varepsilon}_{t(\tau)} F'(w,y^{\varepsilon}(t(\tau+s)))\mu^{\varepsilon}(\tau+s) f_w(w)ds.$$

In the definition of T^{ε}_{τ} and $\hat{A}^{\varepsilon}$, use $F^{\varepsilon}_{t(\tau)}$ just as F^{ε}_{τ} was used in Sections 3 to 5.

Set $f^{\varepsilon}(\tau) = f(w^{\varepsilon}(\tau)) + \varepsilon f_1(w^{\varepsilon}(\tau),\tau)$. Then, it can readily be shown that $f(w^{\varepsilon}(\tau))$ and $f_1(w^{\varepsilon}(\tau),\tau)$ are in the domain of $\hat{A}^{\varepsilon}$ and that

$$\begin{aligned}\hat{A}^{\varepsilon} f^{\varepsilon}(\tau) = {} & f'_w(w^{\varepsilon}(\tau))[G(w^{\varepsilon}(\tau),y^{\varepsilon}(t(\tau)))\mu^{\varepsilon}(\tau) \\ & + F(w^{\varepsilon}(\tau),y^{\varepsilon}(t(\tau)))\frac{\mu^{\varepsilon}(\tau)}{\varepsilon} + H(w^{\varepsilon}(\tau))Y] \\ & - F'(w^{\varepsilon}(\tau),y^{\varepsilon}(t(\tau))f_w(w^{\varepsilon}(\tau))\frac{\mu^{\varepsilon}(\tau)}{\varepsilon} + \varepsilon f^{\varepsilon\prime}_{1,w}(w^{\varepsilon}(\tau),\tau)\dot{w}^{\varepsilon}(\tau),\end{aligned}$$

which equals $A(Y)f(w^{\varepsilon}(\tau)) + O(\varepsilon)$ (we dropped the ℓ_1 subscript on $A(Y)$). This yields convergence of the finite dimensional distributions of $w^{\varepsilon}(\cdot)$ to those of $w(\cdot)$, as in Theorem 2. Tightness is also proved in the same way as done in Theorem 2, completing the proof under (D2). Note that the function f_2, which we used in Theorem 2, is not needed here.

Now, we prove the assertion under (D1) and (C3). In this case $|\dot{w}^{\varepsilon}(\tau)|$ is bounded uniformly in $\varepsilon > 0$ and

$\tau \leq 1$ and $w^{\varepsilon}(0) \to \tilde{x}(0)$ weakly. Thus $\{w^{\varepsilon}(\cdot)\}$ is tight in $C^{m}[0,1]$ and so is the function with values

$$\int_0^t [\frac{\mu^{\varepsilon}(\tau)}{\varepsilon} F(w^{\varepsilon}(\tau),y^{\varepsilon}(t(\tau))) + G(w^{\varepsilon}(\tau),y^{\varepsilon}(t(\tau)))\mu^{\varepsilon}(\tau)]d\tau.$$

Consequently, drawing a convergent subsequence and indexing it also by ε we have that $w^{\varepsilon}(\cdot)$ converges weakly to a continuous process $\bar{w}(\cdot)$. By using a Skorokhod imbedding technique, we can assume for our purposes that the convergence is w.p.1, uniformly on bounded intervals and write

$$\bar{w}(v) = \tilde{x}(0) + \int_0^v H(\bar{w}(s))Yds + \lim_{\varepsilon} \int_0^v \frac{\mu^{\varepsilon}(\tau)}{\varepsilon} F(w^{\varepsilon}(\tau),y^{\varepsilon}(t(\tau)))d\tau. \tag{9.4}$$

We wish to show that the limit in (9.4) iz zero w.p.1 for each v. If so, then since it is continuous w.p.1, it must be identical zero w.p.1. Then, by the uniqueness (C3), $\bar{w}(\cdot) = w(\cdot)$ and the proof will be concluded.

The limit equals

$$\lim_{\varepsilon} \int_0^v \frac{\mu^{\varepsilon}(\tau)}{\varepsilon} F(\bar{w}(\tau),y^{\varepsilon}(t(\tau)))d\tau$$

by the continuity of F and boundedness of $y(\cdot)$. Let $\alpha > 0$. define $\bar{w}^{\alpha}(t) = \bar{w}(m\alpha)$ on $[m\alpha,m\alpha+\alpha)$ for each integer m. The difference between the last limit and $\lim_{\varepsilon} \int_0^v \frac{\mu^{\varepsilon}(\tau)}{\varepsilon} F(\bar{w}^{\alpha}(t),y^{\varepsilon}(t(\tau)))d\tau$ goes to zero as $\alpha \to 0$. Thus, we need only show that $\lim_{\varepsilon} \int_{m\alpha}^{m\alpha+\alpha} \frac{\mu^{\varepsilon}(\tau)}{\varepsilon} F(\bar{w}(m\alpha),y^{\varepsilon}(t(\tau)))d\tau$ is zero w.p.1 for each α. Now, by changing the time scale back to the original one, the last limit equals

$\lim_{\varepsilon} \int_{t(m\alpha)}^{t(m\alpha+\alpha)} \frac{F(\bar{w}(m\alpha), y^{\varepsilon}(u))}{\varepsilon} du$. Since $t(n\alpha) \to 0$ for each n as $\varepsilon \to 0$, the results of Theorem 2 imply that this last limit is zero w.p.1. Q.E.D.

BIBLIOGRAPHY

[1] T.G. Kurtz, "Semigroups of conditional shifts and approximation of Markov processes", Ann. Prob., (1975), 4, 618-642.

[2] R.Z. Khasminskii, "A limit theorem for solutions of differential equations with random right hand sides", Theory of Prob. and Applic., (1966), 11, 390-406.

[3] G.C. Papanicolaou, "Some probabilistic problems and methods in singular perturbations", Rocky Mountain Journal of Math., (1976), 6, 653-674.

[4] G.C. Papanicolaou, and W. Kohler, "Asymptotic theory of mixing stochastic ordinary differential equations", Comm. Pure and Appl. Math., (1974), 27, 641-668.

[5] G. Blankenship, and G.C. Papanicolaou, "Stability and control of stochastic systems with wide-band noise disturbances", SIAM J. on Appl. Math., (1978), 34, 437-476.

[6] E. Wong, and M. Zakai, "On the relationship between ordinary and stochastic differential equations", Int. J. Engin. Science, (1965), 3, 213-229.

[7] P. Billingsley, Convergence of Probability Measures, (1968), John Wiley and Sons, New York.

[8] A.V. Balakrishnan, "Identification of systems subject to random state disturbance", Report UCLA Engin 7348, (1973).

[9] A.V. Balakrishnan, "On the approximation of Itô integrals using band-limited processes", Report UCLA Engin. 7342, (1973).

[10] E.J. McShane, Stochastic Calculus and Stochastic Models (1974), Academic Press, New York.

[11] H.J. Sussmann, "On the gap between deterministic and ordinary differential equations", Ann. Prob. 6, (1978), 19-41.

[12] S.I. Marcus, "Modelling and analysis of stochastic differential equations driven by point processes", IEEE Trans. on Information Theory, (1978), IT-24, 164-172.

[13] I.I. Gikhman, and A.V. Skorokhod, Introduction to the Theory of Random Processes, (1965), Saunders, Philadelphia.

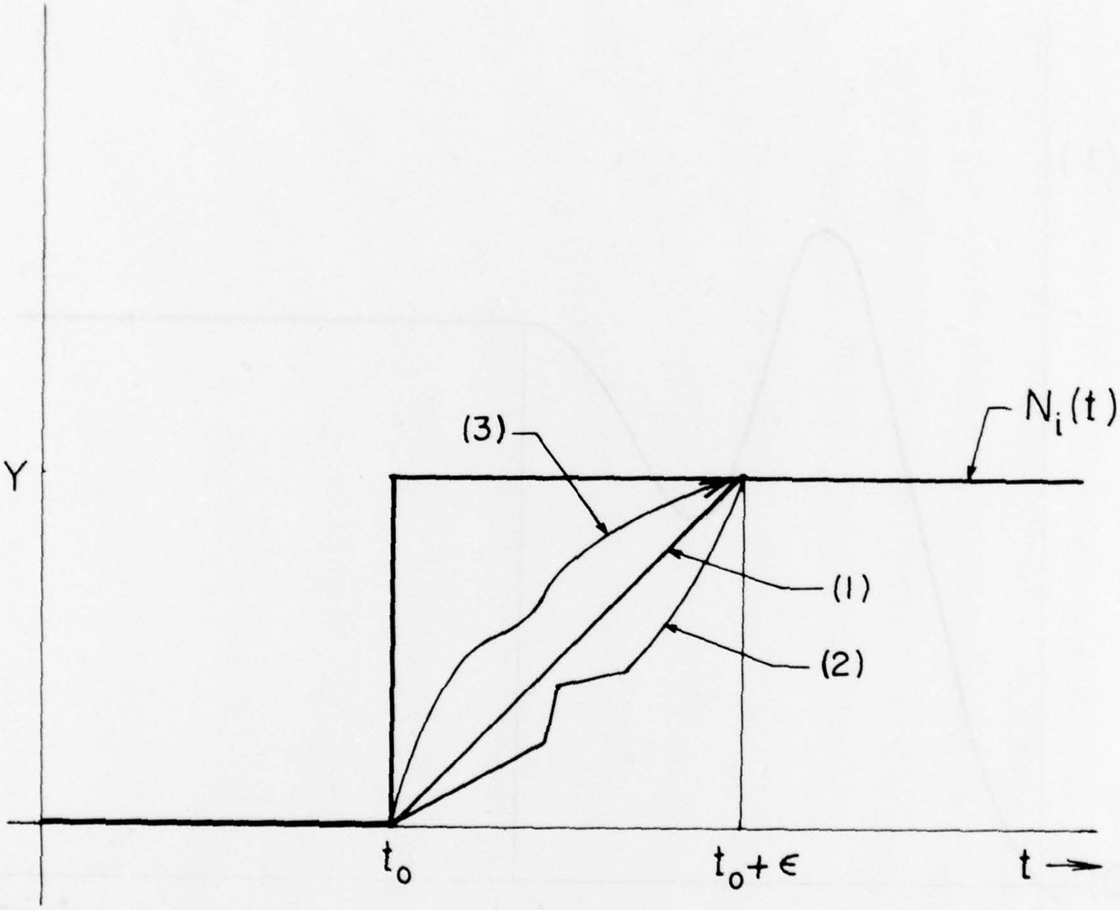


Fig. 1. Illustration of 3 possibilities for $\int_0^t J_i^\epsilon(s)ds$

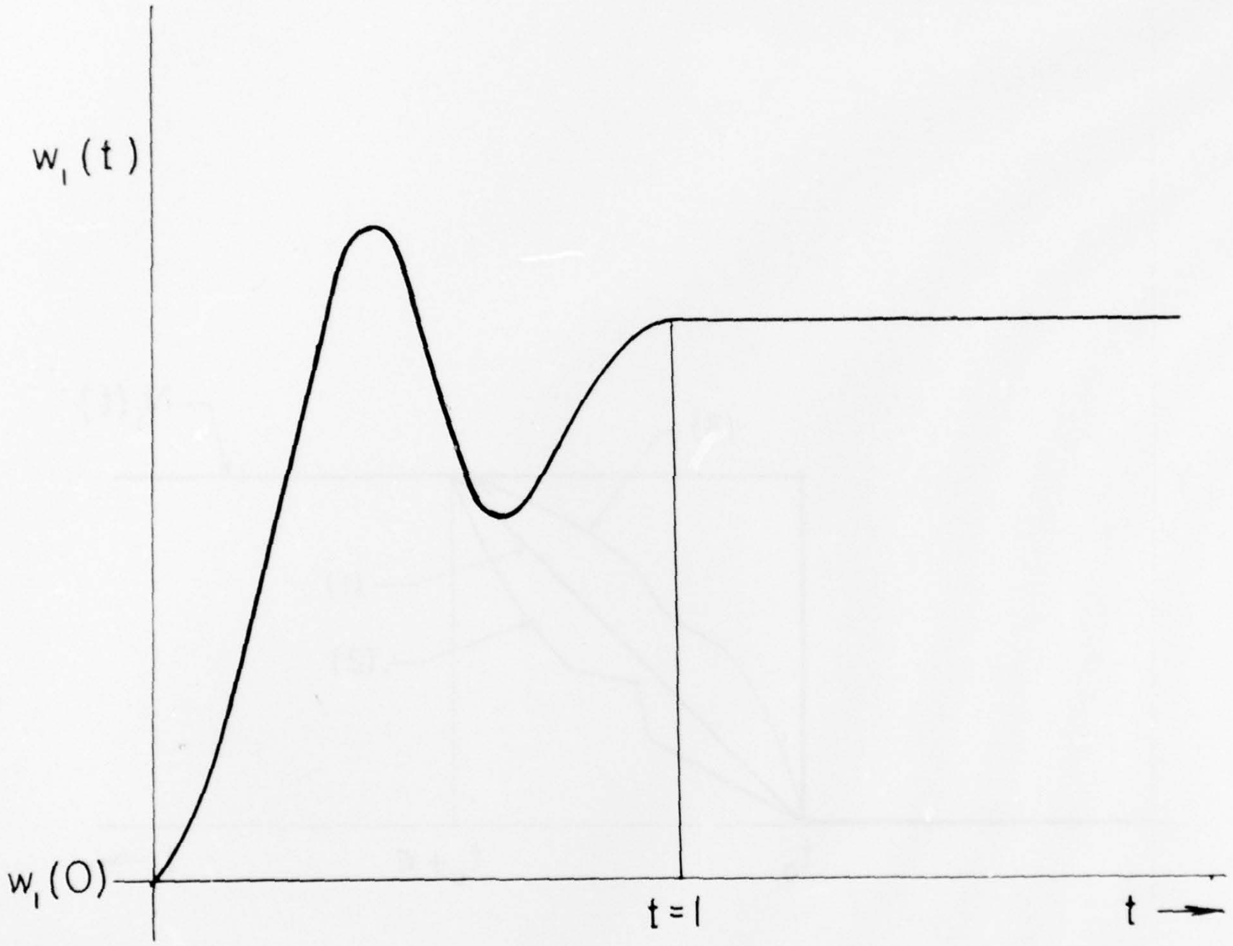


Fig. 2. Possible oscillations of $w_1(t)$ implying lack of weak convergence of $x^\epsilon(\cdot)$